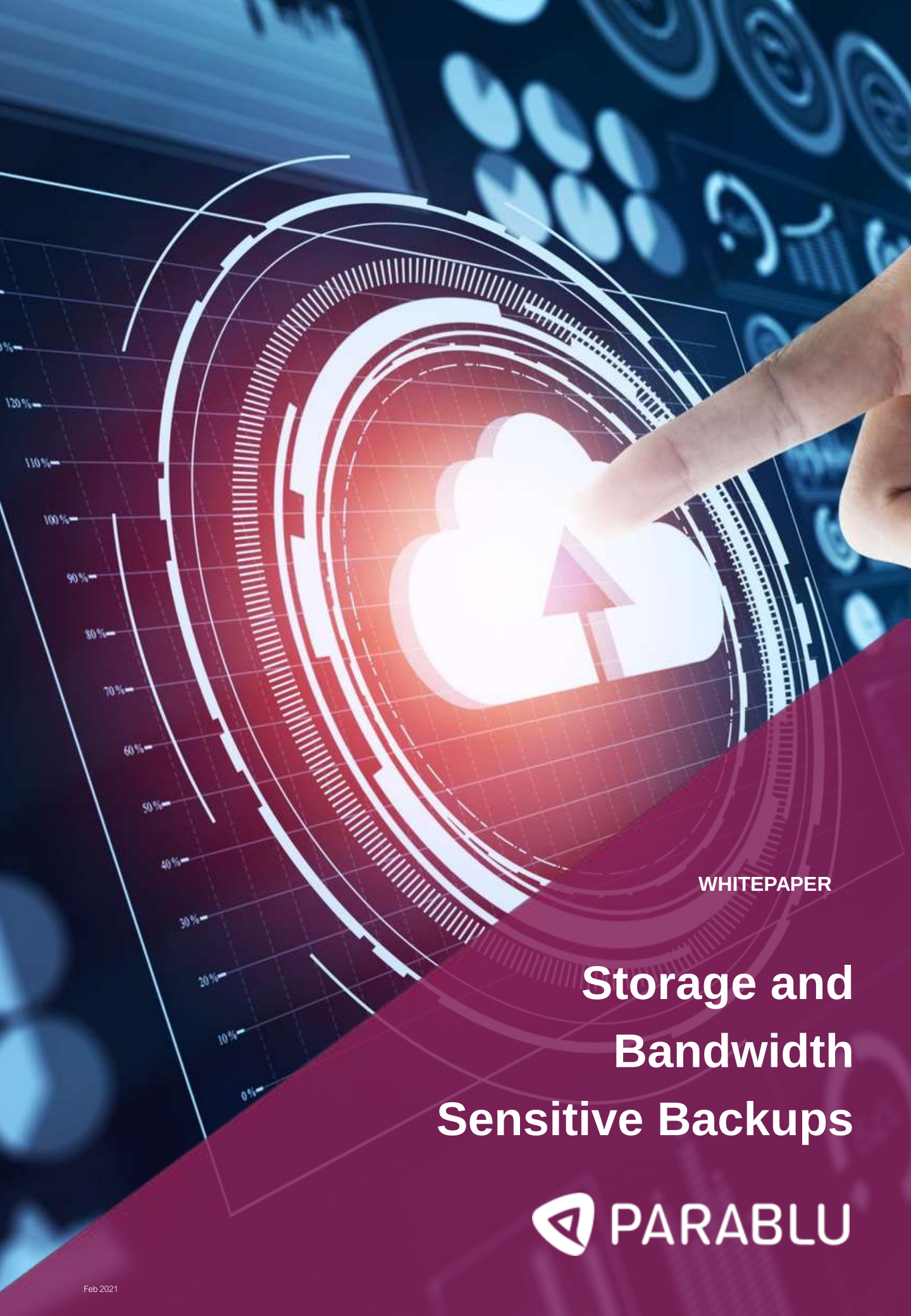


WHITEPAPER

Storage and Bandwidth Sensitive Backups

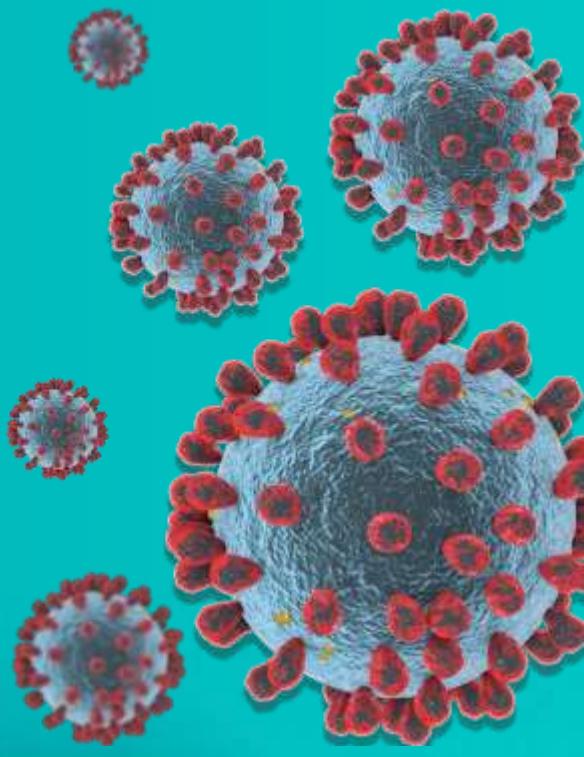
PARABLU

Feb 2021

In the post-pandemic normal

As businesses hurriedly transformed themselves to be ready for the new WFH world, a lot has been written about SaaS, VDI, Zero Trust Security, VPN, MFA, SCCM – all important technologies to keep employees going in the new normal – even as they work from home.

An area that a lot of CIOs and CISOs are now coming around to grappling with is data backup. While data backups were easy (or at least relatively so) pre-pandemic – they've become more daunting with a distributed workforce, working on home networks with slow bandwidths, and questionable reliability. At the same time, with users getting untethered from traditional file servers in the office, and having more of their data on their endpoints, data backup of user data has also become more important than ever. Regulatory compliance, the threat of ransomware, the fear of malicious deletion – are all drivers for businesses to ensure up-to-date protection of endpoints being used for business.



SaaS

Backups can pose a unique challenge when employees work from home. First of all, if you were using an on-premise backup solution that relied on an on-premise backup server attached to secondary storage – that arrangement is unlikely to work well now that your users are at home and unable to reach the backup server. A majority of businesses are opting to use a SaaS-based backup solution that user endpoints can connect, no matter where they are. Customers who already use cloud workloads like or Amazon or Microsoft have moved their backup servers to the cloud as a workload to be run from there. A small number have situated their on-premise servers in their DMZ and allowed inbound HTTPS connections – simulating their own SaaS.

A side-effect of this decision is the impact it has on user authentication. A secure authentication mechanism like Active Directory isn't straightforward when your users are at home. Enter Azure Active Directory, which is a popular Identity Provider alternative. BTW, for those who aren't current with this - Azure Active Directory DOES NOT mean that you need to take your Active Directory server and place it in the cloud – and is just as secure (if not more secure) than traditional AD. So it is certainly worth looking into. Azure Active Directory allows secure user authentication, (SSO) single sign-on, and MFA (multi-factor authentication) allows your users to authenticate reliably and let you sleep peacefully at night.

BANDWIDTH

Once you get past the question of 'how' to perform backups, there is the question of 'how much'. How much data is it practical to backup when your employees are WFH? Do you still back up their entire endpoint? What about large files, like PSTs? Can a home Wi-Fi network handle the load?

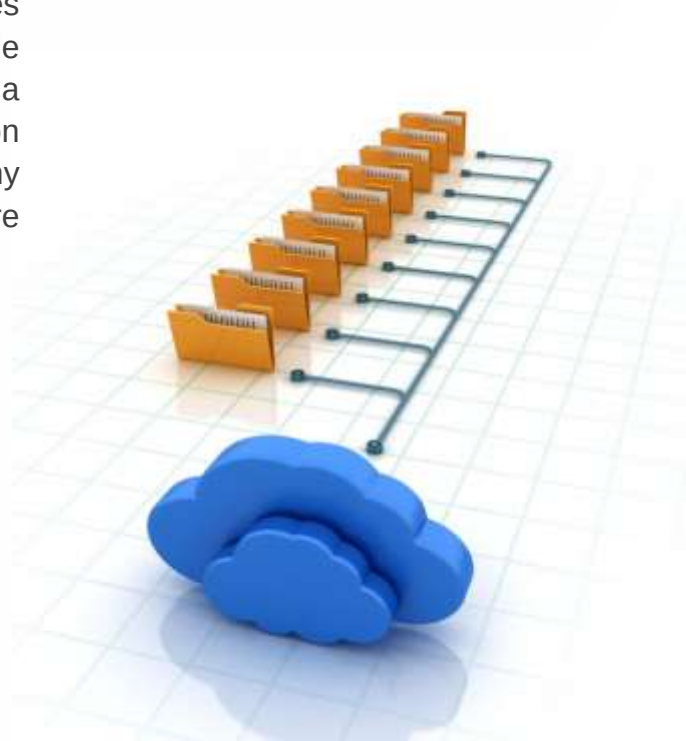
There are a number of techniques that sophisticated, commercial-grade backup solutions use to solve this problem.

INCREMENTAL BACKUPS AND DELTA INCREMENTALS

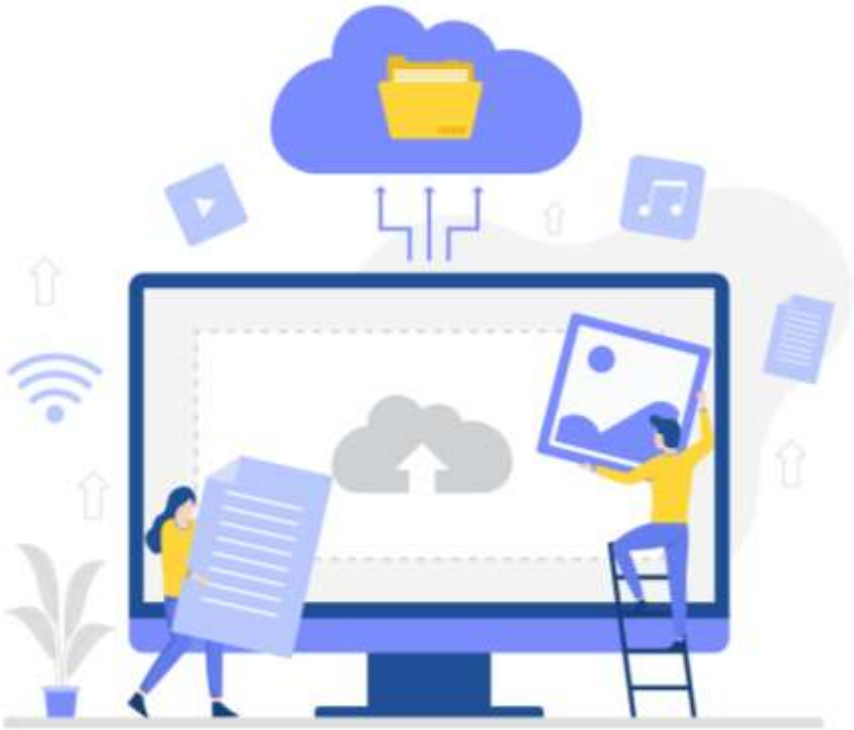
The first, and most obvious is incremental backups. Once there is a full sweep done on the endpoint and an initial copy of all data is made, there is simply no reason to back things up again unless they've changed. An incremental backup identifies files that have changed and only qualifies them for backup. What about restores then, one might ask? How would you keep track of which file got backed up at what time to go retrieve it from the right backup? Fortunately, you won't have to worry about this – most of the time. Good backup software solutions have capable cataloging technologies which can track file changes along a time dimension and bring back all your latest data when you do a restore.

A delta incremental or a partial file backup takes this further. What about very large files (like PSTs for instances) that change only a little bit each day? An incremental backup would qualify the entire PST file for backup each day even only a few emails changed inside it. A delta incremental can identify changes WITHIN files and pick up only changed pieces of data inside a file, for backup. This of course requires a fairly sophisticated cataloging mechanism on the part of the backup software – but many current enterprise-grade software will measure up to this task.

Files like PSTs can also be in-use and 'locked' by an application like Microsoft Outlook, when the backup runs. Conventional sync software (like OneDrive for Business) will simply skip over such files and not back them up. Good backup software designed for enterprise use will handle locked files and manage to backup them up consistently.



BACKUP EVERYTHING OR NOT?



Another good question is whether it is necessary to back up everything off an endpoint. If you have a fairly good idea where your users tend to keep their backup data, you can perhaps identify only such folders to be included in the backup. Or if there are certain folders where your users tend to keep personal data, you could configure the backup to pick up everything on the system – excepting such personal folders. If the backup software allows you to include only certain file types in the backup – take advantage of such a feature. Office documents and PDF files are good starting points for many organizations.

Focus on backing up user data. Don't lug over all the operating system files over a slow network. Operating systems can be re-installed or re-imaged very quickly with today's technologies. Modern backup software should exclude known OS files and folders automatically.

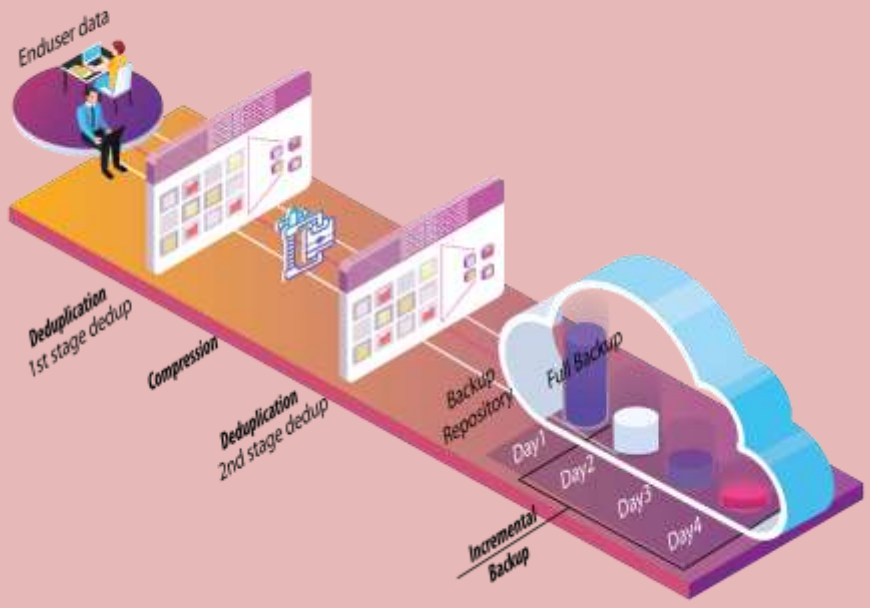
REDUCING NETWORK BANDWIDTH USAGE

There are two popular approaches that backup software solutions take to further reduce network bandwidth (and storage) usage. One is software compression. The other is de-duplication. Both fundamentally approach the problem of space savings from the point of matching identical data fragments. If a fragment of data was already encountered and processed – then its representation is optimized with a redirection to the original copy of the data fragment, thus saving space in the final representation of the sum aggregate of the data.

The efficacy of software compression is usually limited to a single file and is heavily dependent on the type of file. File types that are pre-compressed will not yield much more space even if the backup software attempts to compress it again. De-duplication approaches the problem at a much larger scale by maintaining a persistent database of all known data fragments and optimizing for network and storage usage with a much larger namespace of data signatures or ‘hashes’.

De-duplication can be file-level or block level. File-level de-duplication was popular around 20 years ago and represents early forays into de-duplication technology. One might be fairly hard-pressed to find backup software today that relies only on file-level de-duplication. Any de-duplication claims made today by backup software are likely to be block level. Variable-length block-level de-duplication is fairly state-of-the-art and is what you should be looking for. As to where the de-duplication is happening. The de-duplication should start at the source (i.e. the endpoint itself) and not on the backup server.

Note that de-duplication technology by its very nature, requires highly sophisticated cataloging techniques and there are only a handful of enterprise-class software vendors which provide this reliably.



NETWORK BANDWIDTH THROTTLING

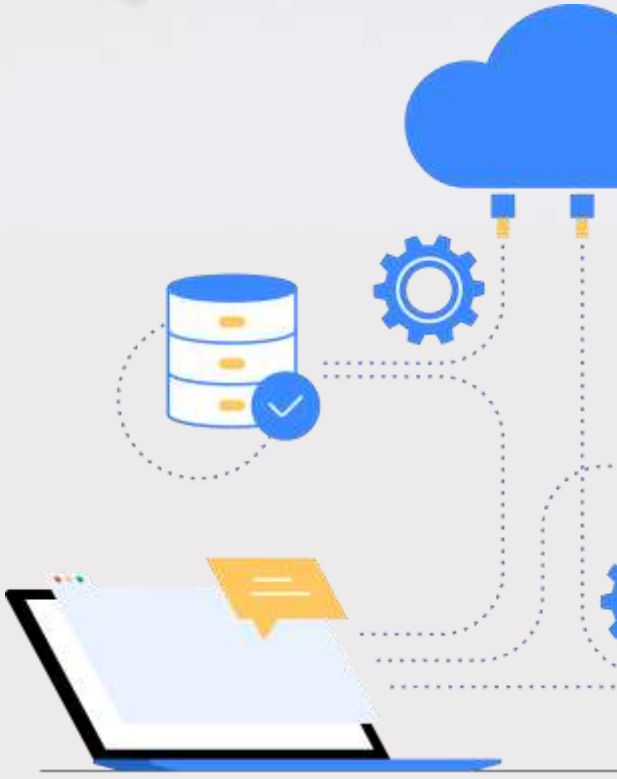
Modern backup software should also give you the option to set up network bandwidth limits to constrain how much bandwidth it can use when running on the user endpoint. This is particularly important in the current climate when a lot of backups aren't happening on leased lines and MPLS links – but over home Wi-Fi networks. The backup software's ability to limit its bandwidth consumption (and CPU consumption) has become an almost necessary element in today's world. The bandwidth may be specified as a percentage or in absolute values (Mbps, Kbps) and an administrator should be able to allow different throttle limits to different groups of users.

STORAGE OPTIMIZATION

A number of the techniques described above also reduce storage footprint as a side effect. When you're transferring fewer data over the network, you're also storing less on the target storage. Storage, however, can form a significant portion of your backup costs. Especially when you consider a meaningful backup payload that is higher than 50GB/user and consider the usage month-on-month and year-on-year. And cloud storage can be expensive.

A good way to optimize storage cost would be to repurpose unused cloud storage allocations you already have. If you are a Microsoft 365 customer, you probably have a ton of OneDrive storage you can re-purpose for backup. We've implemented Parablu's backup solutions this way for several of our customers.

To summarize, there are several ways to overcome the challenge of bandwidth and storage when backing up user data.



PARABLU

info@parablu.com

www.parablu.com

The Cloud is a Great Thing.
We just want to make it safer.